Copula representations and order statistics for conditionally independent random variables

Ismihan Bairamov
Department of Mathematics, Izmir University of Economics, Izmir Turkey

Abstract

The copula representations for conditionally independent random variables and the distribution properties of order statistics of these random variables are studied.

Key words: Copula, Conditional independence, order statistics.

1 Introduction

Dawid (1979) noted that many of the important concepts of statistical theory can be regarded as expressions of conditional independence, thus the conditional independence offers a new language for the expression of statistical concepts and a framework for their study. Dawid (1979) considered the random variables X ,Y and Z. If the random variables X and Y are independent in their joint distribution given $Z = z$, for any value of z, then X and Y are called conditionally independent given Z and denoted by $(X \perp Y \mid Z)$. A general calculus of conditional independence is developed in Dawid (1980), where the general concept of conditional independence for a statistical operation is introduced. Dawid (1980) showed how the conditional independence for statistical operations encompasses the basic properties such as, sufficiency, $(X \perp \Theta \mid T)$, in which X is a random variable with distributions governed by a parameter Θ, and T is a function of X, or.pointwise independence, adequacy etc. Pearl et al. (1989) considered conditional independence as a statement and addressed the problem of representing the sum total of independence statements that logically follow from a given set of such statements. Shaked and Spizzichino (1998) considered n nonnegative random variables $T_i, i = 1, 2, ..., n$ which are interpreted as lifetimes of n units and assuming that $T_1, T_2, ..., T_n$ are conditionally independent given some random variable Θ, determined the conditions under which these conditionally independent random variables are positive dependent. In the Bayesian setting it is of interest to know which kind of dependence arises when Θ is unknown. Prakasa Rao (2006) studied the properties of conditionally independent random variables and proved the conditional versions of generalized Borel-Cantelli

lemma, generalized Kolmogorov's inequality, Hàjek-Rényi inequality. Prakasa-Rao (2006) presented also the conditional versions of classical strong law of large numbers and central limit theorem.

In this paper a different approach to conditionally independent random variables is considered, the necessary and sufficient conditions for conditional independence in terms of the partial derivatives of distribution functions and copulas are given. Also, the distributional properties of order statistics of conditionally independent random variables are studied. The paper is organized as follows: in section 2 we present a definition of conditionally independent random variables $X_1, X_2, ..., X_n$ given Z and derive a sufficient and necessary conditions for conditional independence in terms of copulas. These conditions allow to construct conditionally independent random variables with given bivariate distributions of (X_i, Z) and marginal distributions of X_i's. In Section 3 we study the distributions of order statistics from conditionally independent random variables. It is shown that these distributions can be expressed in terms of partial derivatives of copulas of X_i and Z. The permanent expressions for distributions of order statistics are also presented.

2 Conditionally independent random variables

Let $(X_1, X_2, ..., X_n, Z)$ be $n+1$ variate random vector with joint distribution function (cdf) $H(x_1, x_2, ..., x_n, z) = C(F_{X_1}(x_1), F_{X_2}(x_2), ..., F_{X_n}(x_n), F_Z(z))$, where $F_{X_i}(x_i) = P\{X_i \leq x_i\}$, $i = 1, 2, ..., n$, $F_Z(z) = P\{Z \leq z\}$ and C is a connecting copula.

Definition 1 *If*

$$\begin{aligned} P\{X_1 &\leq x_1, X_2 \leq x_2, ..., X_n \leq x_n \mid Z = z\} = \\ P\{X_1 &\leq x_1 \mid Z = z\} P\{X_2 \leq x_2 \mid Z = z\} \cdots P\{X_n \leq x_n \mid Z = z\} \end{aligned} \quad (1)$$

for all $(x_1, x_2, ..., x_n, z) \in \mathbb{R}^{n+1}$, *then the random variables* $X_1, X_2, ..., X_n$ *are said to be conditional independent given* Z.

It is clear that if the random variables $X_1, X_2, ..., X_n$ are conditionally independent given Z then the conditional random variables $X_{i,z} \equiv (X_i \mid Z = z), i = 1, 2, ..., n$ are independent for all $z \in \mathbb{R}$ and $P\{X_1 \in B_1, ..., X_n \in B_n \mid \mathcal{F}_Z\} = P\{X_1 \in B_1 \mid \mathcal{F}_Z\} \cdots P\{X_n \in B_n \mid \mathcal{F}_Z\}$ a.s. for any Borel sets $B_1, B_2, ..., B_n$, where $\mathcal{F}_Z$ is a $\sigma-$algebra generated by Z.

Example 1 *Bairamov and Arnold (2008) discussed the residual lifelengths of the remaining components in an* $n-k+1-out-of-n$ *system with lifetimes of the components* $X_1, X_2, \ldots, X_n$, *respectively. Let* X_i*' s are independent and identically distributed (iid) random variables with common absolutely continuous distribution F. If we are given* $X_{k:n} = x$, *then the conditional distribution of the subsequent order statistics* $X_{k+1:n}, \ldots, X_{n:k}$ *is the same as the distribution of order statistics of a sample of size* $n-k$ *from the distribution* F *truncated below*

at x. If we denote by $Y_i^{(k)}, i = 1, 2, \ldots, n-k$ the randomly ordered values of $X_{k+1:n}, \ldots, X_{n:n}$, then given $X_{k:n} = x$, these $Y_i^{(k)}$'s will be iid with common survival function $\bar{F}(x+y)/\bar{F}(x)$.

2.1 Assumptions and notations

Throughout this paper we assume that $X_1, X_2, ..., X_n, Z$ are absolutely continuous random variables with corresponding probability density functions (pdf) $f_{X_i}(x), i = 1, 2, ..., n$ and $f_Z(z)$, respectively. Denote by $F_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n)$ and $C_{X_1,X_2,...,X_n}(u_1, ..., u_n)$ the joint distribution function and the copula of $(X_1, X_2, ..., X_n)$, respectively and denote the pdf by $f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n)$. We use the following notations for the bivariate marginal distributions and copulas of random variables X_i and Z :

$$F_{X_i,Z}(x_i, z) \equiv P\{X_i \le x, Z \le z\}, i = 1, 2, ..., n,$$

and

$$F_{X_i,Z}(x_i, z) \equiv C_{X_i,Z}(F_{X_i}(x_i), F_Z(z)),$$

where $C_{X_i,Z}(u, w)$, $(u, w) \in [0, 1]^2$, is the copula of random variables (X_i, Z). We use also the following notations for partial derivatives:

$$\dot{H}(x_1, x_2, ..., x_n, z) = \frac{\partial H(x_1, x_2, ..., x_n, z)}{\partial z}, \ \dot{F}_{X_i,Z}(x_i, z) = \frac{\partial F_{X_i,Z}(x_i, z)}{\partial z},$$

$$\dot{C}(u_1, u_2, ..., u_n, w) \equiv \frac{\partial C(u_1, u_2, ..., u_n, w)}{\partial w}, \ \dot{C}_{X_i,Z}(u, w) \equiv \frac{\partial}{\partial w} C_{X_i,Z}(u, w).$$

The following theorem gives a necessary and sufficient condition for conditional independence of random variables $X_1, X_2, ..., X_n$ given Z.

2.2 Necessary and sufficient conditions for conditional independence

Theorem 1 *If $f_Z(z) > 0$, then the random variables $X_1, X_2, ..., X_n$ are conditionally independent given $Z = z$ if and only if*

$$\dot{H}(x_1, x_2, ..., x_n, z) = \left(\frac{1}{f_Z(z)}\right)^{n-1} \prod_{i=1}^{n} \dot{F}_{X_i,Z}(x_i, z). \tag{2}$$

Proof. It is clear that (1) and (2) both are equivalent to

$$\begin{aligned} \lim_{\Delta z \to 0} P\{X_1 &\le x_1, X_2 \le x_2, ..., X_n \le x_n \mid z \le Z < z + \Delta z\} \\ &= \lim_{\Delta z \to 0} P\{X_1 \le x_1 \mid z \le Z < z + \Delta z\} \cdots P\{X_n \le x_n \mid z \le Z < z + \Delta z\}. \end{aligned}$$

■

Corollary 1 *The random variables $X_1, X_2, ..., X_n$ are conditionally independent given Z if and only if*

$$\dot{C}(u_1, u_2, ..., u_n, w) = \prod_{i=1}^{n} \dot{C}_{X_i,Z}(u_i, w) \text{ for all } 0 \leq u_1, u_2, ..., u_n, w \leq 1. \quad (3)$$

Proof. From (2) one can write

$$\begin{aligned} & \frac{1}{f_Z(z)} \frac{\partial}{\partial z} \left[C(F_{X_1}(x_1), F_{X_2}(x_2), ..., F_{X_n}(x_n), F_Z(z)) \right] \\ = & \left(\frac{1}{f_Z(z)} \right)^n \prod_{i=1}^{n} \frac{\partial}{\partial z} \left[C_{X_i,Z}(F_{X_i}(x_i), F_Z(z)) \right]. \end{aligned} \quad (4)$$

It follows from (4) that

$$\begin{aligned} & \dot{C}(F_{X_1}(x_1), F_{X_2}(x_2), ..., F_{X_n}(x_n), F_Z(z)) \\ = & \prod_{i=1}^{n} \dot{C}_{X_i,Z}(F_{X_i}(x_i), F_Z(z)) \end{aligned}$$

and the transformation $F_{X_i}(x_i) = u_i, (i = 1, 2, ..., n),\ F_Z(z) = w$ leads to (3). ■

Corollary 2 *$X_1, X_2, ..., X_n$ are conditionally independent given Z, if and only if the following integral representations for joint distribution function and copula hold true:*

$$\begin{aligned} F_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) &= \int_{-\infty}^{\infty} \left(\frac{1}{f_Z(z)} \right)^{n-1} \prod_{i=1}^{n} \frac{\partial F_{X_i,Z}(x_i, z)}{\partial z} dz \\ -\infty &< x_1 < \cdots < x_n < \infty \end{aligned} \quad (5)$$

and

$$\begin{aligned} C_{X_1,X_2,...,X_n}(u_1, u_2, ..., u_n) &= \int_{0}^{1} \prod_{i=1}^{n} \dot{C}_{X_i,Z}(u_i, w) dw, \\ 0 &\leq u_1, ..., u_n \leq 1. \end{aligned} \quad (6)$$

Proof. The proof can be made easily by integrating (2) and (3). ■

2.3 Construction of conditionally independent random variables

Using Theorem 1 one can construct the conditionally independent random variables $X_1, X_2, ..., X_n$ given Z, with given joint copulas $C_{X_i,Z}(u_i, w)$, $(u, w) \in [0, 1]^2$, of (X_i, Z), $i = 1, 2, ..., n$. The constructed joint distributions of conditionally independent random variables can be used, for example, in reliability

analysis, for modeling lifetimes of the system having $n+1$ components, such that knowing the exact lifetime of one of the components allows the independence assumption for other components. Consider for example a system of three dependent components with lifetimes X_1, X_2, Z having joint cdf $F_{X_1,X_2,Z}(x_1, x_2, z)$. Assume that the lifetime of the system is $T = \max\{\min(Z, X_1), \min(Z, X_2)\}$ and if $Z = z$ is given, then the interaction between two other components becomes weaker, hence X_1 and X_2 can be assumed to be independent. In this and many similar applications, where the conditional independence is a subject, the constructed joint distributions of conditionally independent random variables can be used for modeling of lifetimes or other random variables of interest. In Shaked and Spizzichino (1998) some interesting applications of conditionally independent random variables, such as an imperfect repair with random effectiveness and lifetimes in random environments are discussed.

The following examples demonstrate the construction procedure of conditionally independent random variables by using Theorem 1.

Example 2 *Let $n = 2$. Denote by $C(u, v, w)$ the copula of random variables (X_1, X_2, Z), by $C_{X_1,Z}(u, w)$ the copula of (X_1, Z) and by $C_{X_2,Z}(v, w)$ the copula of (X_2, Z), respectively. Assume that the joint distributions of (X_1, Z) and (X_2, Z) are classic Farlie-Gumbel-Morgenstern (FGM) distribution, i.e. the copulas are*

$$
\begin{aligned}
C_{X_1,Z}(u, w) &= uw\{1 + \alpha(1 - u)(1 - w)\}, (u, w) \in [0, 1]^2, -1 \le \alpha \le 1 \\
C_{X_2,Z}(v, w) &= vw\{1 + \alpha(1 - v)(1 - w)\}, (v, w) \in [0, 1]^2, -1 \le \alpha \le 1,
\end{aligned}
$$

with

$$
\begin{aligned}
\dot{C}_{X_1,Z}(u, w) &= \frac{\partial}{\partial w}\{uw\{1 + \alpha(1 - u)(1 - v)\}\} \\
&= u + \alpha(1 - u)(1 - 2w), (u, w) \in [0, 1]^2, -1 \le \alpha \le 1 \\
\dot{C}_{X_2,Z}(v, w) &= v + \alpha(1 - v)(1 - 2w), (v, w) \in [0, 1]^2, -1 \le \alpha \le 1.
\end{aligned}
$$

Then from equation (3) we have

$$
\begin{aligned}
\dot{C}(u, v, w) &= \dot{C}_{X_1,Z}(u, w)\dot{C}_{X_2,Z}(v, w) \\
&= [u + \alpha(1 - u)(1 - 2w)]\,[v + \alpha(1 - v)(1 - 2w)] \\
&= uv + \alpha uv(2 - u - v)(1 - 2w) + \alpha^2 uv(1 - u) \qquad (7) \\
&\quad \times(1 - v)(1 - 2w)^2.
\end{aligned}
$$

Integrating (7) with respect to w, one obtains

$$
C(u, v, w) = uvw + \alpha uvw(2 - u - v)(1 - w) - \frac{\alpha^2}{6}uv(1 - u)(1 - v)(1 - 2w)^3. \qquad (8)
$$

Therefore, the random variables X_1, X_2 with the copula

$$
C_{X_1,X_2}(u, v) = uv + \frac{\alpha^2}{6}uv(1 - u)(1 - v)
$$

are conditionally independent if the copula of (X_1, X_2, Z) *is* $C(u, v, w)$ *given in (8).*

Example 3 *Using calculations in Example 1 it is not difficult to see that if the random variables* $X_1, X_2, ..., X_n$ *and* Z *are defined with their copulas as*

$$\begin{aligned} C_{X_1,Z}(u_1, w) &= u_1 w\{1 + \alpha(1 - u_1)(1 - w)\}, \\ C_{X_2,Z}(u_2, w) &= u_2 w\{1 + \alpha(1 - u_2)(1 - w)\}, -1 \le \alpha \le 1 \end{aligned}$$

and

$$C_{X_i,Z}(u_i, w) = u_i w,\ i = 3, 4, ..., n,\ 0 \le u_1, u_2, ..., u_n, w \le 1,$$

then the solution of the equation

$$\dot{C}(u_1, u_2, \cdots u_n, w) = \prod_{i=1}^{n} \dot{C}_{X_i,Z}(u_i, w)$$

is

$$\begin{aligned} C(u_1, u_2, ..., u_n, w) &= u_1 u_2 \cdots u_n w + \alpha u_1 u_2 \cdots u_n w(2 - u_1 - u_2)(1 - w) \\ &\quad - \frac{\alpha^2}{6} u_1 u_2 \cdots u_n (1 - u_1)(1 - u_2)(1 - 2w)^3. \end{aligned} \tag{9}$$

The copula of $X_1, X_2, ..., X_n$ *is*

$$C_{X_1,X_2,...,X_n}(u_1, u_2, ..., u_n) = u_1 u_2 \cdots u_n + \frac{\alpha^2}{6} u_1 u_2 \cdots u_n (1 - u_1)(1 - u_2) \tag{10}$$

and $X_1, X_2, ..., X_n$ *are conditionally independent given* Z.

Remark 1 *The conditional independence of random variables* $X_1, X_2, ..., X_n$ *makes it possible to evaluate many important probabilities concerning dependent random variables by replacing them with the independent pairs of random variables. Assume that we need to calculate the probability of some event connected with the dependent random variables* $X_1, X_2, ..., X_n$*, for example consider*

$$\begin{aligned} P\{(X_1, X_2, ..., X_n) &\in \mathbf{B}\} \\ &= \int_{\mathbf{B}} f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) dx_1 dx_2 \cdots dx_n, \end{aligned}$$

where $\mathbf{B} \in \Re^n$, $\Re^n$ *is the Borel* $\sigma-$*algebra of subsets of* $\mathbb{R}^n$, $\mathbf{B} = B_1 \times B_2 \times \cdots \times B_n$ *and* B_i *are Borel sets on* $\mathbb{R}$ *and* $f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n)$ *is the joint pdf of* $X_1, X_2, ..., X_n$. *If the random variables* $X_1, X_2, ..., X_n$ *are conditionally*

independent given Z, then conditioning on Z we have

$$\begin{aligned}
P\{(X_1, X_2, ..., X_n) &\in B\} = \int_{-\infty}^{\infty} P\{(X_1, X_2, ..., X_n) \in B \mid Z = z\} dF_Z(z) \\
&= \int_{-\infty}^{\infty} \prod_{i=1}^{n} P\{X_i \in B_i \mid Z = z\} dF_Z(z) = \\
&\int_{-\infty}^{\infty} \left[\frac{1}{f_Z(z)}\right]^{n-1} \left\{ \prod_{i=1}^{n} \int_{B_i} \frac{\partial \dot{F}_{X_i,Z}(x_i, z)}{\partial x_i} dx_i \right\} dz. \quad (11)
\end{aligned}$$

Furthermore, if $\mathbf{G}$ is some region on $\mathbb{R}^n$, then

$$\begin{aligned}
P\{(X_1, X_2, ..., X_n) &\in \mathbf{G}\} = \int_{-\infty}^{\infty} P\{(X_1, X_2, ..., X_n) \in \mathbf{G} \mid Z = z\} dF_Z(z) \\
&= \int_{-\infty}^{\infty} P\{(X_{1,z}, X_{2,z}, ..., X_{n,z}) \in \mathbf{G}\} dF_Z(z)
\end{aligned}$$

$$= \int_{-\infty}^{\infty} \left(\int_{\mathbf{G}} f_{X_{1,z}}(x_1) f_{X_{2,z}}(x_2) \cdots f_{X_{n,z}}(x_n) dx_1 dx_2 \cdots dx_n \right) dF_Z(z), \qquad (12)$$

where $X_{i,z} = (X_i \mid Z = z), i = 1, 2, ..., n$ are independent random variables with pdf's

$$f_{X_{i,z}}(x_i) = \frac{1}{f_Z(z)} \frac{\partial \dot{F}_{X_i,Z}(x_i, z)}{\partial x_i}, i = 1, 2, ..., n,$$

respectively. In the following example we illustrate how to use (12) in calculating of stress-strength probability $P\{X_1 < X_2\}$ for dependent, but conditionally independent random variables.

Example 4 *First we construct conditionally independent random variables X_1, X_2 given $Z = z$. Let*

$$\begin{aligned}
F_{X_1,Z}(u, z) &= u^2 z\{1 + (1 - u^2)(1 - z)\},\ F_{X_1}(u) = u^2, F_Z(z) = z,\ 0 \le u, z \le 1 \\
F_{X_2,Z}(v, z) &= vz\{1 + (1 - v)(1 - z)\},\ F_{X_2}(v) = v, F_Z(z) = z,\ 0 \le v, z \le 1,
\end{aligned}$$

It is easy to calculate

$$\begin{aligned}
\dot{F}_{X_1,Z}(u, z) &= u^2[1 - (1 - u^2)(1 - z)] + u^2 z(1 - u^2) \qquad (13) \\
\dot{F}_{X_2,Z}(v, z) &= v[1 + (1 - v)(1 - z)] + vz(v - 1). \qquad (14)
\end{aligned}$$

Then using Theorem 1 we have

$$\dot{F}_{X_1,X_2,Z}(u, v, z) = \dot{F}_{X_1,Z}(u, z) \dot{F}_{X_2,Z}(v, z) dz. \qquad (15)$$

Using (13), (14 in 15) and then integrating we have

$$\begin{aligned} & F_{X_1,X_2,Z}(u,v,z) \\ = \;& \frac{4}{3}u^2v^2z^3 - \frac{4}{3}u^2vz^3 + \frac{4}{3}u^4vz^3 - \frac{4}{3}u^4v^2z^3 - 3z^2u^4v + 2z^2u^4v^2 + 2z^2u^2v \\ & -z^2u^2v^2 + 2u^4vz - u^4v^2z. \end{aligned} \tag{16}$$

The joint cdf of X_1 and X_2 is

$$F_{X_1,X_2}(u,v) = \frac{2}{3}u^2v + \frac{1}{3}u^2v^2 + \frac{1}{3}u^4v - \frac{1}{3}u^4v^2 \tag{17}$$

and the pdf is

$$f_{X_1,X_2}(u,v) = \frac{4}{3}u + \frac{4}{3}uv + \frac{4}{3}u^3 - \frac{8}{3}u^3v.$$

The corresponding copula of (X_1, X_2) can be obtained by using transformation $F_{X_1}(u) = u^2 = t$, $F_{X_2}(v) = v = s$ and it is

$$\begin{aligned} C_{X_1,X_2}(t,s) &= \frac{2}{3}ts + \frac{1}{3}ts^2 + \frac{1}{3}t^2s - \frac{1}{3}t^2s^2, \\ 0 &\leq t,s \leq 1. \end{aligned}$$

It follows from Theorem 1 that X_1, X_2 are conditionally independent given $Z = z$.

Now consider the probability $P\{X_1 < X_2\}$. By usual way integrating $f_{X_1,X_2}(u,v)$ over the set $\{(u,v) : u < v\}$ we obtain

$$P\{X_1 < X_2\} = \int_0^1 \int_0^v f_{X_1,X_2}(u,v)dudv = \frac{31}{90}. \tag{18}$$

On the other hand using conditional independence of X_1, X_2 given Z, we have

$$\begin{aligned} P\{X_1 &< X_2\} = \int P\{X_1 < X_2 \mid Z = z\}dF_Z(z) \\ &= \int P\{X_{1,z} < X_{2,z}\}dF_Z(z) \\ &= \int_0^1 \left(\int_0^1 F_{1,z}(u)dF_{2,z}(u) \right) dF_Z(z), \end{aligned} \tag{19}$$

where $X_{1,z} \equiv X_1 \mid Z = z$ and $X_{2,z} \equiv X_2 \mid Z = z$ are independent random variables with cdf's

$$\begin{aligned} F_{X_{1,z}}(x) &= \frac{1}{f_Z(z)}\dot{F}_{X_1,Z}(x,z) \\ \text{and } F_{X_{2,z}}(x) &= \frac{1}{f_Z(z)}\dot{F}_{X_2,Z}(x,z), \end{aligned}$$

respectively. Since $f_Z(z)=1$, then from (19) taking into account (13) and (14) we have

$$\begin{aligned}
P\{X_1 &< X_2\} = \int_0^1 \left(\int_0^1 \left\{ u^2[1-(1-u^2)(1-z)] + u^2 z(1-u^2) \right\} dF_{2,z}(u) \right) dz \\
&= \int_0^1 \left(\int_0^1 \left\{ u^2[1-(1-u^2)(1-z)] + u^2 z(1-u^2) \right\} \{1+(1-2u)(1-2z)\} \, du \right) dz \\
&= \int_0^1 \left(\frac{1}{15} + \frac{7}{15} z + \frac{2}{15} z^2 \right) dz = \frac{31}{90}, \qquad (20)
\end{aligned}$$

which agrees with (18).

3 Order statistics

Denote by $X_{1:n}, X_{2:n}, ..., X_{n:n}$ the order statistics of conditionally independent random variables $X_1, X_2, ..., X_n$ given Z. We are interested in distribution of order statistics $X_{r:n}$, $1 \leq r \leq n$ and the joint distributions of $X_{r:n}$ and $X_{s:n}$, $1 \leq r < s \leq n$. The formulas for distributions of order statistics contains expressions depending on copulas of pairs (X_i, Z), $i = 1, 2, ..., n$ and permanents.

3.1 Distribution of a single order statistics

Conditioning on Z one obtains

$$\begin{aligned}
P\{X_{n:n} &\leq x\} = \int_{-\infty}^{\infty} P\{X_{n:n} \leq x \mid Z = z\} dF_Z(z) \\
&= \int_{-\infty}^{\infty} \prod_{i=1}^n P\{X_i \leq x \mid Z = z\} dF_Z(z) = \int_{-\infty}^{\infty} \left(\frac{1}{f_Z(z)} \right)^n \prod_{i=1}^n \frac{\partial F_{X_i,Z}(x_i, z)}{\partial z} dF_Z(z) \\
&= \int_{-\infty}^{\infty} \prod_{i=1}^n \dot{C}_{X_i,Z}(F_i(x_i), F_Z(z)) dF_Z(z) = \int_0^1 \prod_{i=1}^n \dot{C}_{X_i,Z}(F_i(x_i), s) ds
\end{aligned}$$

and similarly

$$P\{X_{1:n} \leq x\} = 1 - \int_0^1 \prod_{i=1}^n [1 - \dot{C}_{X_i,Z}(F_i(x_i), s)] ds.$$

The distribution of $X_{r:n}$,$1 \leq x \leq n$ can be derived as follows:

$$\begin{aligned}
F_{r:n}(x) &\equiv P\{X_{r:n} \leq x\} = \sum_{i=r}^{n} \int_{-\infty}^{\infty} P\{\text{exactly } i \text{ of } X\text{'}s \text{ are } \leq x \mid Z = z\} dF_Z(z) \\
&= \sum_{i=r}^{n} \frac{1}{i!(n-i)!} \sum_{(j_1, j_2, \ldots, j_n) \in S_n} \int_{-\infty}^{\infty} \left\{ \prod_{l=1}^{i} P\{X_{j_l} \leq x \mid Z = z\} \right. \\
&\quad \times \left. \prod_{l=i+1}^{n} [1 - P\{X_{j_l} \leq x \mid Z = z\}] \right\} dF_Z(z) \\
&= \sum_{i=r}^{n} \frac{1}{i!(n-i)!} \sum_{(j_1, j_2, \ldots, j_n) \in S_n} \int_{0}^{1} \left\{ \dot{C}_{X_{j_l}, Z}(F_{X_{j_l}}(x), s) \right. \qquad (21) \\
&\quad \times \left. \prod_{l=i+1}^{n} \left[1 - \dot{C}_{X_{j_l}, Z}(F_{X_{j_l}}(x), s)\right] \right\} ds,
\end{aligned}$$

where

$$S_n = \{(j_1, j_2, ..., j_n), 1 \leq j_1, j_2, ..., j_n \leq n\}, \qquad (22)$$

is the set of all $n!$ permutations of $(1, 2, ..., n)$ and $\sum_{(j_1, j_2, \ldots, j_n) \in S_n}$ extends over all elements of S_n.

Remark 2 *If $X_1, X_2, ..., X_n$ are identically distributed, i.e. $F_{X_i}(x) = F_X(x), \forall x \in \mathbb{R}, i = 1, 2, ..., n$ and the joint distributions of $(X_1, Z), (X_2, Z), ..., (X_n, Z)$ are the same, i.e. $C_{X_i, Z}(u, w) = C_{X,Z}(u, w), \forall (u, w) \in [0, 1]^2$, $i = 1, 2, ..., n$ then*

$$F_{r:n}(x) = \sum_{i=r}^{n} \binom{n}{i} \int_{0}^{1} \left[\dot{C}_{X,Z}(F_X(x), s)\right]^i \left[1 - \dot{C}_{X,Z}(F_X(x), s)\right]^{n-i} ds. \qquad (23)$$

Remark 3 *It follows from (23) that, if $C_{X_i, Z}(u, w) = uw$, i.e. $X_1, X_2, ..., X_n$ are independent and identically distributed (iid) random variables, then $\dot{C}_{X,Z}(F_X(x), s) = F_X(x)$ and*

$$F_{r:n}(x) = \sum_{i=r}^{n} \binom{n}{i} F_X^i(x) \left[1 - F_X(x)\right]^{n-i}.$$

3.2 The joint distributions of two order statistics

The joint distribution of $X_{r:n}$, and $X_{s:n}$ is

$$F_{r,s}(x, y) = P\{X_{r:n} \leq x, X_{s:n} \leq y\}$$

$$
\begin{aligned}
= & \sum_{i=r}^{n} \sum_{j=\max(s-i,0)}^{n-i} \frac{1}{i!(j-i)!(n-i-j)!} \int_{-\infty}^{\infty} \left\{ \sum_{(l_1,\ldots,l_n)\in S_n} \prod_{k=1}^{i} P\{X_{l_k} \le x \mid Z = z\} \right. \\
& \times \prod_{k=i+1}^{i+j} [P\{X_{l_k} \le y \mid Z = z\} - P\{X_{l_k} \le x \mid Z = z\}] \\
& \left. \times \prod_{k=i+j+1}^{n} [1 - P\{X_{l_k} \le y \mid Z = z\}] \right\} dF_Z(z)
\end{aligned}
$$

and in terms of copulas

$$F_{r,s}(x,y)$$

$$
\begin{aligned}
= & \sum_{i=r}^{n} \sum_{j=\max(s-i,0)}^{n-i} \frac{1}{i!(j-i)!(n-i-j)!} \sum_{(l_1,\ldots,l_n)\in S_n} \int_{0}^{1} \left\{ \prod_{k=1}^{i} \dot{C}_{X_{l_k},Z}(F_{X_{l_k}}(x), s) \right. \\
& \times \prod_{k=i+1}^{i+j} \left[\dot{C}_{X_{l_k},Z}(F_{X_{l_k}}(y), s) - \dot{C}_{X_{l_k},Z}(F_{l_k}(x), s) \right] \\
& \left. \times \prod_{k=i+j+1}^{n} \left[1 - \dot{C}_{X_{l_k},Z}(F_{X_{l_k}}(y), s) \right] \right\} ds \qquad (24)
\end{aligned}
$$

Remark 4 *If $X_1, X_2, ..., X_n$ are identically distributed with cdf $F_X(x)$, and the joint distributions of $(X_1, Z), (X_2, Z), ..., (X_n, Z)$ are the same, i.e. $C_{X_i,Z}(u,w) = C_{X,Z}(u,w), \forall (u,w) \in [0,1]^2$, $i = 1, 2, ..., n$, then*

$$
\begin{aligned}
F_{r,s}(x,y) = & \sum_{i=r}^{n} \sum_{j=\max(s-i,0)}^{n-i} \frac{n!}{i!(j-i)!(n-i-j)!} \int_{0}^{1} \left[\dot{C}_{X,Z}(F_X(x), s) \right]^{i} \\
& \times \left[\dot{C}_{X,Z}(F_X(y), s) - \dot{C}_{X_{l_k},Z}(F_X(x), s) \right]^{j-i} \qquad (25) \\
& \times \left[1 - \dot{C}_{X,Z}(F_X(y), s) \right]^{n-i-j} ds
\end{aligned}
$$

Remark 5 *It follows from (25) that, if $C_{X_i,Z}(u,w) = uw$, i.e. $X_1, X_2, ..., X_n$ are iid random variables, then $\dot{C}_{X,Z}(F_X(x), s) = F_X(x)$ and*

$$
\begin{aligned}
F_{r,s}(x,y) = & \sum_{i=r}^{n} \sum_{j=\max(s-i,0)}^{n-i} \frac{n!}{i!(j-i)!(n-i-j)!} F_X^i(x) \\
& \times (F_X(y) - F_X(x))^{j-i} \left[1 - F_X(y) \right]^{n-i-j},
\end{aligned}
$$

which agrees the well known formula for joint cdf of rth and sth order statistics (see David and Nagaraja (2003))

3.3 Expressions for joint distributions of order statistics with permanents

Suppose $A = (a_{ij}), i, j = 1, 2, ..., n$ is the square matrix. The permanent of A is defined as

$$Per(A) = \sum_{(j_1, j_2, \dots, j_n) \in S_n} \prod_{k=1}^{n} a_{k, j_k},$$

where S_n is defined in (22) and $\sum_{(j_1, j_2, \dots, j_n) \in S_n}$ denotes the sum over all $n!$permutations $(j_1, j_2, ..., j_n)$ of $(1, 2, ..., n)$. Using (21) one can realize that

$$F_{r:n}(x) = \sum_{i=r}^{n} \frac{1}{i!(n-i)!} \int_0^1 Per\left(M_1(x, s)\right) ds,$$

where

$$M_1(x,s)$$
$$= \begin{array}{c} \left(\begin{array}{cccc} \dot{C}_{X_1,Z}(F_{X_1}(x), s) & \dot{C}_{X_2,Z}(F_{X_2}(x), s) & \cdots & \dot{C}_{X_n,Z}(F_{X_n}(x), s) \\ \dot{C}_{X_1,Z}(F_{X_1}(x), x) & \dot{C}_{X_2,Z}(F_{X_2}(x), s) & \cdots & \dot{C}_{X_n,Z}(F_{X_n}(x), s) \\ \vdots & \vdots & & \vdots \\ \dot{C}_{X_1,Z}(F_{X_1}(x), s) & \dot{C}_{X_2,Z}(F_{X_2}(x), s) & \cdots & \dot{C}_{X_n,Z}(F_{X_n}(x), s) \\ & & & \\ 1-\dot{C}_{X_1,Z}(F_{X_1}(x), s) & 1-\dot{C}_{X_2,Z}(F_{X_2}(x), s) & \cdots & 1-\dot{C}_{X_n,Z}(F_{X_n}(x), s) \\ 1-\dot{C}_{X_1,Z}(F_{X_1}(x), s) & 1-\dot{C}_{X_2,Z}(F_{X_2}(x), s) & \cdots & 1-\dot{C}_{X_n,Z}(F_{X_n}(x), s) \\ \vdots & \vdots & & \vdots \\ 1-\dot{C}_{X_1,Z}(F_{X_1}(x), s) & 1-\dot{C}_{X_2,Z}(F_{X_2}(x), s) & \cdots & 1-\dot{C}_{X_n,Z}(F_{X_n}(x), s) \end{array} \right) \end{array} \begin{array}{l} \left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} \begin{array}{l} i \\ \text{times} \end{array} \\ \\ \left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} \begin{array}{l} n-i \\ \text{times} \end{array} \end{array} .$$

Similar permanent expression for joint distribution of order statistics can be obtained from (24)

$$F_{r,s}(x, y) = \sum_{i=r}^{n} \sum_{j=\max(s-i,0)}^{n-i} \frac{1}{i!(j-i)!(n-i-j)!} \int_0^1 M_2(x, y, s) ds,$$

where

$$M_2(x, y, s) =$$
$$= \left(\begin{array}{cccc} \dot{C}_{X_1,Z}(F_{X_1}(x), s) & \cdots & \cdots & \dot{C}_{X_n,Z}(F_{X_n}(x), s) \\ & & & \\ \begin{array}{c} \dot{C}_{X_1,Z}(F_{X_1}(y), s) \\ -\dot{C}_{X_1,Z}(F_{X_1}(x), s) \end{array} & \cdots & \cdots & \begin{array}{c} \dot{C}_{X_n,Z}(F_{X_n}(y), s) \\ -\dot{C}_{X_n,Z}(F_{X_n}(x), s) \end{array} \\ & & & \\ 1-\dot{C}_{X_1,Z}(F_{X_1}(x), s) & \cdots & \cdots & 1-\dot{C}_{X_n,Z}(F_{X_n}(x), s) \end{array} \right) \begin{array}{l} \}i \\ \\ \}j-i \\ \\ \}n-i-j \end{array}$$

In general the joint distribution function of order statistics $X_{r_1:n}, X_{r_2:n}, ..., X_{r_k:n}$, $1 \leq r_1 < r_2 < \cdots < r_k \leq n$ is

$$F_{r_1,r_2...,r_k}(x_1, x_2, ..., x_k) = \sum \frac{1}{j_1!j_2!\cdots j_k j_{k+1}!} \int\limits_0^1 M_3(x_1, x_2, ..., x_k, s)ds,$$

$$-\infty \quad < \quad x_1 < x_2 < \cdots < x_k < \infty,$$

where the sum is over $j_1, j_2, ..., j_{k+1}$ with $j_1 \geq r_1, j_1 + j_2 \geq r_2, ..., j_1 + j_2 + \cdots + j_k \geq r_k$ and $j_1 + j_2 + \cdots + j_k + j_{k+1} = n$ and

$$M_3(x_1, x_2, ..., x_k, s)$$
$$= \begin{array}{c} \left(\begin{array}{ccc} \dot{C}_{X_1,Z}(F_{X_1}(x_1), s) & \cdots\ \cdots & \dot{C}_{X_n,Z}(F_{X_n}(x_1), s) \\ \begin{array}{c}\dot{C}_{X_1,Z}(F_{X_1}(x_2), s) \\ -\dot{C}_{X_1,Z}(F_{X_1}(x_1), s)\end{array} & \cdots\ \cdots & \begin{array}{c}\dot{C}_{X_n,Z}(F_{X_n}(x_2), s) \\ -\dot{C}_{X_n,Z}(F_{X_n}(x_1), s)\end{array} \\ \vdots & & \vdots \\ \begin{array}{c}\dot{C}_{X_1,Z}(F_{X_1}(x_k), s) \\ -\dot{C}_{X_1,Z}(F_{X_1}(x_{k-1}), s)\end{array} & \cdots\ \cdots & \begin{array}{c}\dot{C}_{X_n,Z}(F_{X_n}(x_k), s) \\ -\dot{C}_{X_n,Z}(F_{X_n}(x_{k-1}), s)\end{array} \\ 1 - \dot{C}_{X_1,Z}(F_{X_1}(x_k), s) & \cdots\ \cdots & 1 - \dot{C}_{X_n,Z}(F_{X_n}(x_k), s) \end{array} \right) \end{array} \begin{array}{l} \}j_1 \\ \\ \}j_2 \\ \\ \}j_k \\ \\ \}j_{k+1} \end{array} .$$

Remark 6 *If $X_1, X_2, ..., X_n$ are identically distributed with cdf $F_X(x)$, and the joint distributions of $(X_1, Z), (X_2, Z), ..., (X_n, Z)$ are the same, i.e. $C_{X_i,Z}(u, w) = C_{X,Z}(u, w), \forall (u, w) \in [0, 1]^2$, $i = 1, 2, ..., n$, then*

$$F_{r_1,r_2,...,r_k}(x_1, x_2, ..., x_n)$$
$$= \sum \frac{1}{j_1!j_2!\cdots j_k j_{k+1}!} \int\limits_0^1 \left[\dot{C}_{X,Z}(F_X(x_1), s)\right]^{j_1} \left[\dot{C}_{X,Z}(F_X(x_2), s) - \dot{C}_{X,Z}(F_X(x_1), s)\right]^{j_2}$$
$$\times \left[\dot{C}_{X,Z}(F_X(x_k), s) - \dot{C}_{X,Z}(F_X(x_{k-1}), s)\right]^{j_k} \left[1 - \dot{C}_{X,Z}(F_X(x_k), s)\right]^{j^{k+1}} ds,$$
$$-\infty \quad < \quad x_1 < x_2 < \cdots < x_k < \infty,$$

where the sum is over $j_1, j_2, ..., j_{k+1}$ with $j_1 \geq r_1, j_1 + j_2 \geq r_2, ..., j_1 + j_2 + \cdots + j_k \geq r_k$ and $j_1 + j_2 + \cdots + j_k + j_{k+1} = n$.

The order statistics are widely used in statistical theory of reliability. The coherent system consisting of $n-$ components with lifetimes $X_1, X_2, ..., X_n$ has lifetime which can be expressed with the order statistics (or the linear functions of order statistics using Samaniego's signatures). The $n - k + 1$-*out-of-n* coherent system, for example, has lifetime $T = X_{k:n}$ and the mean residual life function of such a system at the system level is $\Psi(t) = E\{X_{k:n} - t \mid X_{r:n} > t\}$, $r < k$. The function $\Psi(t)$ expresses the mean residual life (MRL) length of a $n - k + 1$-*out-of-n* system given that at least $n - r + 1$ components are alive

at the moment t. It is clear that to compute the reliability or the MRL functions of such systems we need the joint distributions of order statistics. The results presented in this paper can be used if the system has components with conditionally independent lifetimes.

References

[1] Bairamov, I. and Arnold, B.C. (2008) On the residual lifelengths of the remaining components in an n - k + 1 out of n system. *Statist. Probab. Letters.* 78, 945-952.

[2] David, H.A. and Nagaraja, H.N. (2003) Order Statistics, Third Edition, Wiley, New York.

[3] Dawid, P.A (1979) Conditional independence in statistical theory. *J.R. Statist.Soc. B*, 41, No. 1, pp. 1-31.

[4] Dawid, P.A. (1980) Conditional independence for statistical operations. *Ann. Statist.*, 8, No.3, pp. 598-617.

[5] Pearl J., Geiger D. and Verma, S. T. (1989) Conditional independence and its representations. *Kybernetika*, 25, No. Suppl, pp. 33-34.

[6] Prakasa Rao, B.L.S. (2009) Conditional independence, conditional mixing and conditional association. *Ann Inst Stat Math*, 61, 441-460.

[7] Shaked, M. and Spizzichino, F. (1998) Positive dependence properties of conditionally independent random lifetimes. *Mathematics of Operations Research*, 23, No.4, pp.944-959.